BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 9: e181043
doi: 10.3897/biss.9.181043

OPEN ACCESS

Conference Abstract

Darwin Core Data Package: A Practical Evolution to Support Richer, Deeper, and New Biodiversity Data Sharing

John Richard Wieczorek‡, Tim Robertson§, Paula F Zermoglio|

‡ Rauthiflor LLC, Bariloche, Argentina
§ Global Biodiversity Information Facility, Copenhagen, Denmark
| Universidad Nacional de Río Negro, CONICET, Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural, Bariloche, Argentina

Corresponding author: John Richard Wieczorek (gtuco.btuco@gmail.com)

Received: 02 Dec 2025 | Published: 05 Dec 2025

Citation: Wieczorek J, Robertson T, Zermoglio P (2025) Darwin Core Data Package: A Practical Evolution to Support Richer, Deeper, and New Biodiversity Data Sharing. Biodiversity Information Science and Standards 9: e181043. https://doi.org/10.3897/biss.9.181043

Abstract

After more than fifteen years of use, the **Darwin Core Archive (DwC-A)** (GBIF 2021) has become the most popular means of sharing species occurrence and sampling-event data. But it has a limitation: in its “star-schema” structure, all information must connect directly to a single “core” table. As biodiversity data publishing expands to include molecular, ecological, and behavioral information, publishers increasingly encounter cases where valuable relationships among entities—events, agents, organisms, materials, media—cannot be expressed, resulting in the loss of information.

The **Darwin Core Data Package (DwC-DP)** (Darwin Core Maintenance Group 2025a) arose from this tension. It represents a practical evolution of Darwin Core standard (Wieczorek et al. 2012), retaining its familiar vocabulary while adopting a more flexible, relational structure. Rather than replacing DwC-A, DwC-DP complements it, offering an alternative for datasets that challenge DwC-A’s constraints. DwC-DP supports nested hierarchies of MaterialEntities, Media, and Occurrences in addition to the hierarchical Events already possible in DwC-A.

© Wieczorek J et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DwC-DP was developed in parallel with and implements a proposed **Darwin Core Conceptual Model**, (Darwin Core Maintenance Group 2025b) which provides the semantic foundation for the relationships it supports. The proposed **Darwin Core Data Package Guide**, under review starting 2025-09-17, embodies specifications for using the open Frictionless Data Package framework (Fowler et al. 2017). The Global Biodiversity Information Facility (GBIF) led iterative prototyping and testing beginning in July 2021. The two documents went under public review as potential additions to the Darwin Core standard in late 2025.

One might ask, "Why not simply modify the **Darwin Core Text Guide** specification (Darwin Core Maintenance Group 2023), on which DwC-A is based, to encompass DwC-DP's capabilities?" If the DwC-A specification were changed to rely on Data Package technology alone, existing Darwin Core Archives would become obsolete. That would violate the principle that users satisfied with DwC-As should be able to continue to use them unimpeded. Conversely, a dual specification preserving backward compatibility would not be interpretable by standard tools. By supporting both, we gain the advantage of broader compatibility and integration across disciplines that data packages offer as products of a wider data-science community.

The presentation (Suppl. material 1) from which this extended abstract derives introduces DwC-DP as a story about expressiveness. Using two contrasting narratives—a "DwC-A detective short story" and a "DwC-DP epic novel"—it illustrates how richer relational connections can transform how biodiversity data are documented and the stories they can tell. The "detective short story" is a metaphor for DwC-A's thematic constraints: it follows a single bee, observed, photographed, collected, and identified—an entirely solvable case using DwC-A. The "epic novel" elaborates on a pollination-survey network, in which the survey target consists of bees visiting flowers (organism interactions), material samples yield DNA sequences to aid identification, and observations, media, identifications, and measurements interlink across multiple events and materials. DwC-DP can fully capture this intricate web of relationships. DwC-A cannot.

The use cases that inspired DwC-DP depend on one or more kinds of Events (Occurrence Events, Survey Events, Collecting Events, OrganismInteraction Events), but DwC-DP has no concept of a "core" table as DwC-A does. Instead, Darwin Core Events, Occurrences, and MaterialEntities function as "foundation classes," to which additional "modules" may be attached for richer information: agent roles, assertions, bibliographic resources, chronometric ages, geological contexts, identifiers, identifications, media, nucleotide analyses, organisms, organism interactions, protocols, provenance, surveys, and usage policies. Each module contributes an additional layer of narrative complexity to the "epic novel."

While offering great potential for richness and depth, DwC-DP's modular schemas and simplification mechanisms (e.g., embedded agent identifiers for common roles) let publishers include only what they need—an abridged version of the "epic novel." DwC-DP resembles DwC-A, both use tabular data files and metadata in Ecological Markup Language (EML; Fegraus et al. 2005). The key difference is that DwC-DP describes its

structure in a JSON file (*datapackage.json*), whereas DwC-A uses an XML file (*meta.xml*). The similarity provides a bridge for those familiar with DwC-A who wish to explore DwC-DP's greater expressiveness.

Who should be interested in DwC-DP? Data publishers who struggle to fit complex data into DwC-A; researchers seeking complete, reproducible, citable datasets with their full context preserved; policy and decision-makers relying on diverse primary biodiversity data; and data aggregators who need to support any of the above.

Keywords

DwC-DP, biodiversity data, data standard, data sharing

Presenting author

John Richard Wieczorek

Presented at

Living Data 2025

Funding program

Global Biodiversity Information Facility contract to John Wieczorek, Rauthiflor LLC.

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Darwin Core Maintenance Group (2023) Darwin Core text guide. http://rs.tdwg.org/dwc/terms/guides/text/2023-09-13. Accessed on: 2025-10-24.
- Darwin Core Maintenance Group (2025a) Darwin Core Data Package guide. http://rs.tdwg.org/dwc/doc/dp/2025-09-10. Accessed on: 2025-10-24.
- Darwin Core Maintenance Group (2025b) Darwin Core conceptual model. https://gbif.github.io/dwc-dp/cm/. Accessed on: 2025-10-24.
- Fegraus E, Andelman S, Jones MB, Schildhauer M (2005) Maximizing the Value of Ecological Data with Structured Metadata: An Introduction to Ecological Metadata Language (EML) and Principles for Metadata Creation. Bulletin of the Ecological Society of America 86 (3): 158 -168 . https://doi.org/10.1890/0012-9623(2005)86[158:MTVOED]2.0.CO;2

- Fowler D, Barratt J, Walsh P (2017) Frictionless Data: Making research data quality visible. International Journal of Digital Curation 12 (2): 274-285. https://doi.org/10.2218/ijdc.v12i2.577
- GBIF (2021) Darwin Core Archives - How-to Guide, version 2.2. GBIF Secretariat, Copenhagen. URL: https://ipt.gbif.org/manual/en/ipt/3.0/dwca-guide
- Wieczorek J, Bloom D, Guralnick R, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7 (1): 29715. https://doi.org/10.1371/journal.pone.0029715

Supplementary material

Suppl. material 1: Darwin Core Data Package: A Practical Evolution to Support Richer, Deeper, and New Biodiversity Data Sharing doi

Authors: John Wieczorek, Tim Robertson, Paula Zermoglio
Data type: presentation
Brief description: The original presentation given at Living Data 2025, formatted as a PDF, with speaker notes.
Download file (3.32 MB)